

Refereed paper

Construction and validation of a scoring system for the selection of high-quality data in a Spanish population primary care database (SIDIAP)

M^a del Mar García-Gil

Senior Researcher SIDIAP Database, Institut d'Investigació en Atenció Primària (IDIAP Jordi Gol), Catalonia, Spain and Medical Science Department, Universitat de Girona, Spain

Eduardo Hermosilla

Statistician. SIDIAP Database, Institut d'Investigació en Atenció Primària (IDIAP Jordi Gol), Catalonia, Spain

Daniel Prieto-Alhambra

Scientific Coordinator SIDIAP Database, Institut d'Investigació en Atenció Primària (IDIAP Jordi Gol), Catalonia, Spain, URFOA, IMIM (Institut de Recerca Hospital del Mar). Barcelona, Spain, Institut Català de la Salut, Catalonia, Spain and Universitat Autònoma de Barcelona, Bellaterra, Spain

Francesc Fina

Technic Coordinator SIDIAP Database

Magdalena Rosell

Physician, Scientific Division

Institut d'Investigació en Atenció Primària (IDIAP Jordi Gol), Catalonia, Spain and Institut Català de la Salut, Catalonia, Spain

Rafel Ramos

Senior Researcher SIDIAP Database, Institut d'Investigació en Atenció Primària (IDIAP Jordi Gol), Catalonia, Spain, Medical Science Department, Universitat de Girona, Spain, Institut Català de la Salut, Catalonia, Spain and Institut d'Investigació Biomèdica de Girona, Girona, Spain

Jordi Rodriguez

SIDIAP Database Manager, Institut d'Investigació en Atenció Primària (IDIAP Jordi Gol), Catalonia, Spain

Tim Williams

Research Team Manager, General Practice Research Database (GPRD)

Tjeerd Van Staa

Head of Research, General Practice Research Database (GPRD)

Bonaventura Bolívar

Scientific Director, Institut d'Investigació en Atenció Primària (IDIAP Jordi Gol), Catalonia, Spain

ABSTRACT

Background Computerised databases of primary care clinical records are widely used for epidemiological research. In Catalonia, the Information System for the Development of Research in Primary Care (SIDIAP) aims to promote the development of research based on high-quality validated data from primary care electronic medical records.

Objective The purpose of this study is to create and validate a scoring system (Registry Quality Score, RQS) that will enable all primary care practices (PCPs) to be selected as providers of research-usable data based on the completeness of their registers.

Methods Diseases that were likely to be representative of common diagnoses seen in primary care were selected for RQS calculations. The observed/expected cases ratio was calculated for each disease. Once we had obtained an estimated value for this ratio for each of the selected conditions we added up the ratios calculated for each condition to obtain a final RQS. Rate comparisons between observed

and published prevalences of diseases not included in the RQS calculations (atrial fibrillation, diabetes, obesity, schizophrenia, stroke, urinary incontinence and Crohn's disease) were used to set the RQS cut-off which will enable researchers to select PCPs with research-usable data.

Results Apart from Crohn's disease, all prevalences were the same as those published from the RQS fourth quintile (60th percentile) onwards. This RQS cut-off provided a total population of 1 936 443 (39.6% of the total SIDIAP population).

Conclusions SIDIAP is highly representative of the population of Catalonia in terms of geographical, age and sex distributions. We report the usefulness of rate comparison as a valid method to establish research-usable data within primary care electronic medical records.

Keywords: database management systems, medical records, primary health care, registers, validation studies

What was already known

- Primary care databases, containing validated data coded in electronic medical records provide a powerful source of data for epidemiological research.
- Several methods have been used to assess the completeness and accuracy of registers in such data.

What this study added to our knowledge

- We report, for the first time, the usefulness of rate comparison as a valid method for establishing research-usable data within primary care electronic medical records.
- We also introduce SIDIAP to the scientific community. SIDIAP is one of the few primary care databases containing information on Southern European populations.

Introduction

Computerised databases of primary care clinical records are widely used for epidemiological research, particularly in studies of disease prevalence and incidence, studies of health services and in pharmacoepidemiological research.¹ In Catalonia, the Information System for the Development of Research in Primary Care (SIDIAP) was created in 2010 by the Catalan Institute of Health (ICS) and the Jordi Gol Primary Care Research Institute (IDIAP Jordi Gol). Its main aim is to promote the development of research based on high-quality validated data from primary care electronic medical records.² SIDIAP contains anonymised longitudinal patient information including sociodemographic characteristics, morbidity (Inter-

national Classification of Diseases; ICD-10), clinical and lifestyle variables, laboratory tests and treatments (drug prescriptions, drugs purchased at the community pharmacy and hospital discharge information).

However, data from electronic primary care records are collected for clinical practice rather than research purposes and so investigators need to consider not only the validity and completeness of the data contained therein, but also the extent to which this data can be generalised to the population as a whole. There are many similar research databases in Europe, such as the General Practice Research Database (GPRD),³ the MediPlus database⁴ and the Doctors Independent Network database (DIN),⁵ and in the USA,⁶ that are widely used for observational studies. In the majority of these databases, data are entered on a voluntary basis by general practitioners (GPs), who are required

to record prescribing and relevant patient-encounter events in accordance with strict quality standards. Furthermore, data are routinely validated by an ‘up-to-standard’ audit, confirming the quality of data recording in several key areas.¹ By contrast, SIDIAP consists of all the available clinical information from the general population. Given this situation, it is important to develop stringent posterior validation systems of the quality of data in order to adapt them to the specific needs of research.

This study aims to create and validate a scoring system, the Registry Quality Score (RQS), enabling all primary care practices (PCPs) to be selected as providers of research-usable data based on the completeness of their registers.

Methods

Study design

The study was cross-sectional and population-based.

Setting

The primary care structure in the region of Catalonia (north-east Spain) comprises 358 PCPs composed of health professionals and support staff who are responsible for the health care of the population in a given geographical area.

The Catalan Institute of Health manages 274 PCPs; the remainder are managed by other healthcare providers.

PCPs are constituted by three or more basic care units (BCUs), each of which is made up of one GP and one nurse who share a common list of patients. SIDIAP comprises the clinical information coded in the corresponding medical records of all PCPs, with a total of 3414 BCUs. The global adult population assigned to any of these BCUs is 4 859 725 (from 2005 to 2009, 80% of the total population of Catalonia).

Population

BCUs with fewer than 500 people assigned to them were excluded from the analysis with the result that 3310 BCUs were finally included, serving a population of 4 828 792. BCUs with fewer than 500 people assigned to them are typically either created in response to temporary population increases (e.g. in the tourist season) or to specifically enable GPs who perform administrative tasks (e.g. PCP managers and teaching coordinators) to have a lighter workload. The last-year user population (those who were seen by their GP/nurse at least once in the last year) was chosen for setting the RQS cut-off and comprised 3 403 324 people (70%).

Figure 1 shows the criteria for the population selection.

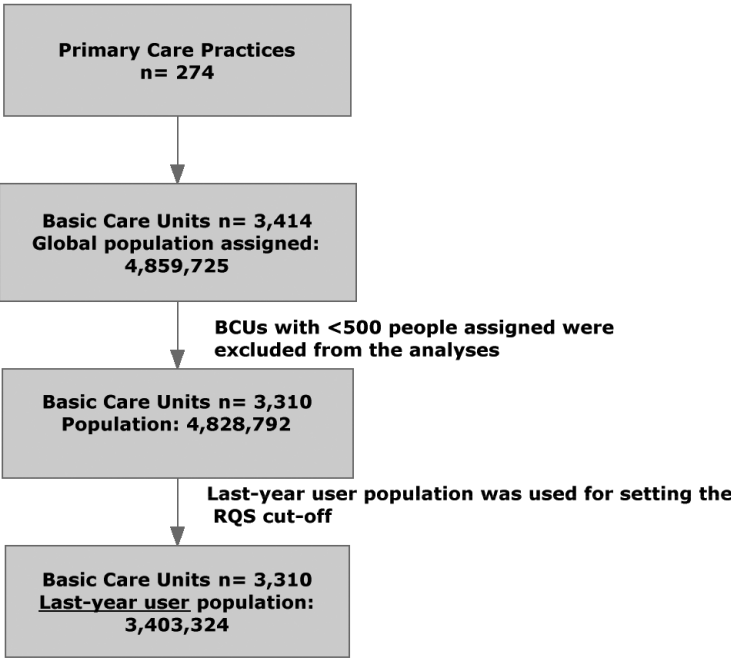


Figure 1 Basic care units and population of the SIDIAP database

RQS calculations

Diseases that were likely to be representative of common diagnoses seen in primary care were selected for RQS calculations. Both pathologies that are used as indicators in evaluating the quality of the health care provided by each GP and those that are not were taken into consideration. The chronic conditions selected were hypertension, chronic obstructive pulmonary disease, heart failure, ischaemic heart disease, osteoarthritis, arthritis and hypothyroidism. Pneumonia and cystitis were included as acute diseases. All of these diseases were ascertained using ICD-10 codes.

The observed/expected cases ratio was calculated for each disease. In the case of chronic conditions, observed cases were the number of people with any of the listed diseases up to 31 December 2009, whereas in the case of acute conditions, the observed cases were the number of people with either of the two diseases newly coded at any point between 1 January 2009 and 31 December 2009. The expected value of diseases by age and sex was defined as the mean value of the prevalence of a disease from all BCUs and was obtained by means of indirect standardisation using the total population as a reference and the specific rates of the conditions by age and sex for each BCU (Box 1).

Once we had obtained an estimated value for this ratio for each one of the selected conditions we added up the ratios calculated for each condition in order to obtain a final score, which we defined as the Registry Quality Score (RQS). Every BCU is assigned with its resulting RQS.

We compared observed and expected prevalences (as published in the available literature) for a list of reference conditions, different from those included in RQS calculations. This ratio was used to set the RQS cut-off, which will allow us to select PCPs as providers of research-usable data. The criteria for selecting the pathologies were the same as those used for the RQS

calculations: long-term and acute conditions often seen in primary care were considered as eligible. The reference conditions finally selected were: atrial fibrillation, diabetes, obesity, schizophrenia, stroke, urinary incontinence and Crohn's disease. Local or high-quality and representative population were the criteria for considering published prevalences in the available literature in order to obtain a reference prevalence/incidence of each of these conditions to which we could compare our estimators.

Statistical analysis

Mean prevalences and their corresponding 95% confidence intervals by specific age and sex distributions of the conditions of reference were calculated according to RQS quintiles. The RQS cut-off was set as the quintile where most of the prevalences were the same as those described in the literature (interval estimation).

For validation purposes, comparison between the total SIDIAP population and the resulting RQS population was then performed in terms of age, sex and the mean prevalences of the diseases used in the RQS calculations. Distribution of the conditions of reference by age and sex were also calculated.

In order to assess the representativeness of the RQS population, the age and sex distribution of the population of Catalonia (2009 census) and the resulting RQS population were compared using a population pyramid plot. Moreover, the participating PCPs (as based on RQS scores for each of their GPs) were represented spatially throughout the territory in order to assess their representativeness.

Analyses were performed using the Statistical Package for the Social Sciences (SPSS), version 13.0, Stata Statistical Software (Stata), release 9, and ArcView 3.2.

Box 1 Standardisation

A principal role in epidemiology is to compare the incidence or prevalence of disease or mortality between two or more populations. However, the comparison of crude mortality or morbidity rates is often misleading because the populations being compared may differ significantly with respect to certain underlying characteristics, such as age or sex, that will affect the overall rate of morbidity or mortality.

One method of overcoming the effects of confounding variables such as age is to combine category-specific rates into a single summary rate that has been adjusted to take into account its age structure or other confounding factor. This is achieved by using the methods of standardisation.

There are two methods of standardisation and these are characterised by whether the standard used is a population distribution (direct method) or a set of specific rates (indirect method). Both direct and indirect standardisation involve the calculation of numbers of expected events (e.g. prevalence), which are compared with the number of observed events.

Results

RQS cut-off

Table 1 shows the mean prevalence of the diseases used in rate comparisons in accordance with the RQS score quintiles. In relation to interval estimation, atrial fibrillation and diabetes prevalence were the same as the literature from the first quintile, whereas the reference for obesity, schizophrenia and stroke corresponded with the second quintile. Urinary incontinence reached the reference interval from the fourth quintile and only Crohn’s disease always showed a lower prevalence rate than the reference. Hence, apart from Crohn’s disease, all prevalences are the same as the reference from the fourth quintile (60th percentile) onwards. This RQS cut-off provides a total population available of 1 936 443 (39.6% of the total SIDIAP population).

RQS validation

RQS general characteristics

Table 2 shows that the RQS population is similar to the SIDIAP population with respect to age and sex distribution. However, the mean prevalence of the diseases

used for the RQS scoring are, as expected, slightly higher in the RQS population.

Prevalences for conditions used to validate RQS by age and sex

As seen in Figure 2, prevalence rates increase gradually with age for atrial fibrillation, stroke and diabetes in both sexes, although these prevalences are somewhat greater in men than in women. Urinary incontinence also increases with age but remains more prevalent in women. With regards to obesity, a steep rise is observed from about 30 years of age in both sexes, although this is more marked in women, and a peak is reached between 50 and 70 years. Finally, schizophrenia and Crohn’s disease appear to be more prevalent at younger ages. Schizophrenia is more frequent in men, whereas no differences in prevalence between sexes are observed in the case of Crohn’s disease.

RQS population structure and geographical representativeness

Figure 3a shows the comparison between the RQS age–sex population and the population of Catalonia (census of 2009) and Figure 3b shows the geographical distribution of the existing 274 PCPs in Catalonia.

Table 1 Rate comparison. RQS cut-off (1-year user population; n = 3 403 324)

	Conditions of reference (age range)						
	AF (> 40 years)	Diabetes (35–74 years)	Obesity (25–60 years)	Schizo- phrenia (15–54 years)	Stroke (35–79 years)	UI (women > 65 years)	Crohn’s disease (all ages)
RQS quintiles							
First	2.37 (2.32–2.41)	7.67 (7.59–7.75)	8.57 (8.48–8.67)	0.68 (0.65–0.70)	1.72 (1.68–1.76)	6.66 (6.50–6.82)	0.10 (0.09–0.11)
Second	2.82 (2.77–2.87)	8.21 (8.13–8.29)	10.52 (10.42–10.61)	0.74 (0.72–0.77)	1.99 (1.95–2.03)	8.90 (8.70–9.11)	0.11 (0.10–0.12)
Third	2.85 (2.80–2.89)	8.49 (8.41–8.57)	11.14 (11.04–11.24)	0.76 (0.74–0.79)	2.09 (2.04–2.13)	9.21 (9.03–9.39)	0.11 (0.11–0.12)
Fourth	2.92 (2.87–2.97)	8.66 (8.58–8.74)	11.87 (11.77–11.97)	0.77 (0.75–0.80)	2.15 (2.11–2.19)	9.93 (9.74–10.12)	0.12 (0.12–0.13)
Fifth	3.00 (2.96–3.05)	9.24 (9.16–9.32)	13.53 (13.42–13.63)	0.85 (0.82–0.88)	2.28 (2.24–2.33)	11.47 (11.27–11.68)	0.13 (0.12–0.14)
Reference rates ^b	2.52 (1.58–4.01)	7.0 (6.7–7.4)	11.2 (10.10– 12.3)	0.80 (0.73–0.88)	2.24 (1.90–2.63)	10–20 ^a	0.18 (0.15–0.21)

Notes: AF, atrial fibrillation; UI, urinary incontinence. ^a Range. ^b See refs 12–18.

Table 2 Sociodemographic characteristics and RQS internal prevalences. Comparison between SIDIAP population and RQS population

	SIDIAP (<i>n</i> = 4 828 792)	RQS population (<i>n</i> = 1 936 443)
Mean age years (SD)	46.19 (18.73)	45.59 (18.85)
Women	50.70 (50.65–50.74)	50.71 (50.64–50.78)
Hypertension	17.74 (17.70–17.77)	18.75 (18.70–18.81)
COPD	2.32 (2.30–2.33)	8.40 (8.34–8.45)
> 65 years	2.70 (2.68–2.72)	10.56 (10.45–10.66)
Ischaemic heart disease	2.26 (2.25–2.27)	2.45 (2.43–2.47)
Heart failure	0.90 (0.90–0.91)	4.13 4.09–4.17)
> 65 years	1.09 (1.08–1.11)	5.70 (5.62–5.78)
Rheumatoid arthritis	0.37 (0.37–0.38)	0.45 (0.44–0.46)
Hypothyroidism	1.67 (1.66–1.68)	1.99 (1.97–2.01)
Cystitis	0.67 (0.67–0.68)	1.09 (1.08–1.10)
Pneumonia	0.32 (0.32–0.33)	0.44 (0.43–0.45)
Osteoarthritis	0.68 (0.67–0.69)	2.18 (2.15–2.21)
> 65 years	0.89 (0.87–0.90)	3.17 (3.11–3.23)

Notes: Values are given as percentages and 95% CI unless otherwise stated. COPD, chronic obstructive pulmonary disease.

PCPs from most of the territory have been included in the RQS. Black dots represents PCPs where at least one BCU is included in the RQS and white dots represent PCPs without any BCU in the RQS.

Discussion

Summary of the main findings

SIDIAP comprises most of the clinical information recorded by primary care health professionals (GPs and nurses) and administrative staff in electronic medical records. The database contains this information for almost five million people, representing approximately 80% of the total population aged over 15 years old in the region of Catalonia (north-east Spain).

We report here the methods used to create and validate a scoring system (RQS) that can be used to choose BCUs with a good quality of coding, as defined by the completeness of the registers. As shown, 40% of the participating professionals with the highest RQS score achieve, for all of the long-term and acute conditions explored except Crohn’s disease, preva-

lence and incidence rates that are comparable with those published in the available literature. Hence, we propose to use the 60th RQS percentile as a suitable cut-off to establish what can be defined as research-usable information. Using this cut-off, we can provide reliable clinical data on about two million people, and on a total of almost ten million person-years for the period 2005–2009.

The RQS score for each BCU will be updated on a six-monthly basis, and data corresponding to the up-to-date RQS will be used to decide which participants will be excluded.

Comparison with existing literature

Rate comparison is a widely used method for the validation of several variables in primary care databases and has been used in many publications to explore the completeness of the information contained in well-known databases such as the GPRD^{7,8}. Rate comparison has also been used as a method to assess the quality of coding of some particular conditions in the same database (e.g. chickenpox, hay fever, asthma and diabetes⁹) and to validate similar sources of information for monitoring certain prescriptions.¹⁰

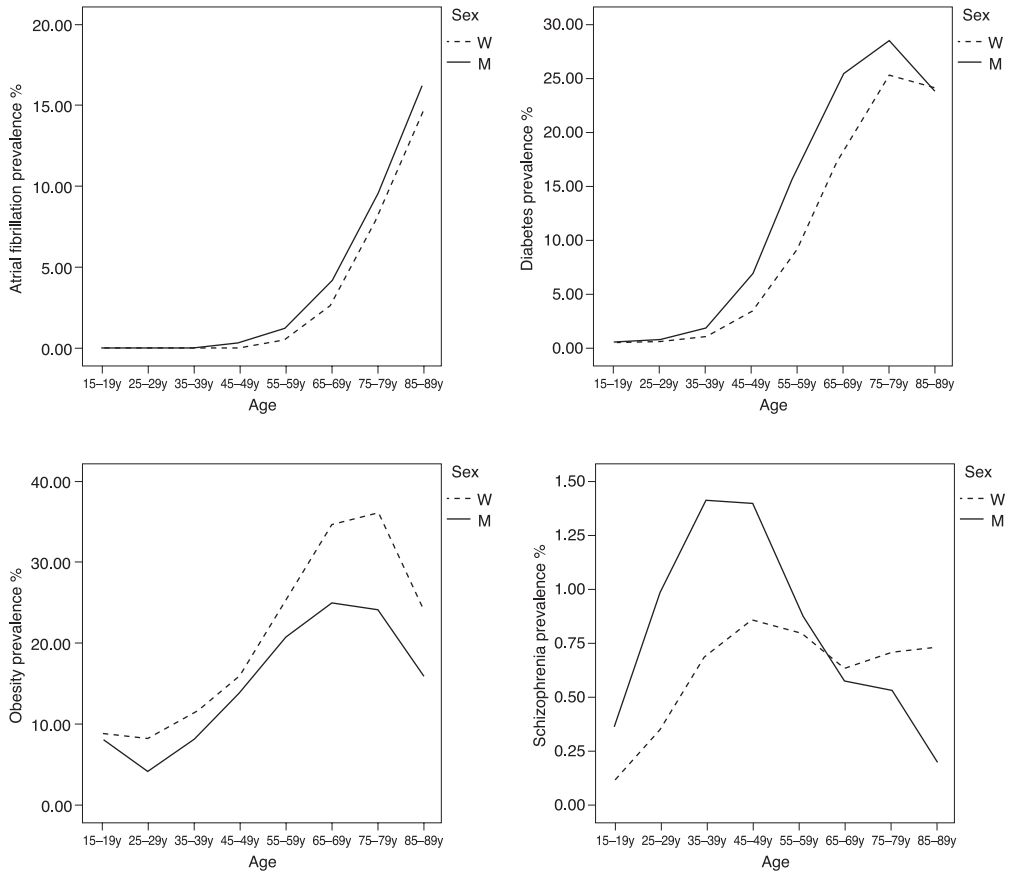


Figure 2 RQS reference prevalences by age and sex

However, we propose the use of a newly created score – RQS – based on rate comparison to decide on the usefulness of data for research purposes.

Alternative methods using parameters such as checks in the continuity of data; relative rates of recording for various events including referrals, prescribing and immunisations; logical recording practices such as the recording of indications with acute prescribing and appropriate recording of registration details; and checks on mortality rates as a proxy for mass deletions of old patients, have all been used to select research-usable data in similar primary care databases, such as the DIN database,⁵ and the GPRD.¹ However, none of these methods has proven to provide better data quality than the others.

In the particular case of our database, previous studies¹¹ have shown that the main problem with the data is lack of completeness due to the under-recording of certain conditions and thus our proposed method was considered, a priori, to be a useful tool in identifying health professionals with a good quality of coding and research-usable data. Furthermore, these methods are consistent with the nature of our database, which does not consist of information sent by volunteer participants (like GPRD or QResearch) but

of the whole set of PCPs in the Institut Català de la Salut (Catalan health service). By using the rate comparison method with the RQS score, permitting the identification of research-usable data, allowance is made for the fact that GPs entering data as part of their standard clinical practice may not have the same level of awareness and motivation as volunteer participants. Further validation studies may show these methods to be generalisable to other similar databases.

As is seen in the results, our findings show that the RQS, which is based on the prevalence/incidence rates of nine specific conditions, correlates well with the reference rates for atrial fibrillation,¹² diabetes,¹³ obesity,¹⁴ schizophrenia,¹⁵ stroke¹⁶ and urinary incontinence.¹⁷ In the case of Crohn's disease,¹⁸ the finding of a lower prevalence than expected may be explained by the fact that most of the prevalence studies reviewed used screening methods that are not applicable to a general population attending primary care.

We have also shown that when age- and sex-specific prevalences are considered in the RQS population, all of the studied conditions fit with their known epidemiological pattern.^{12–18} This gives further support to the good degree of accuracy of the coding in this population.

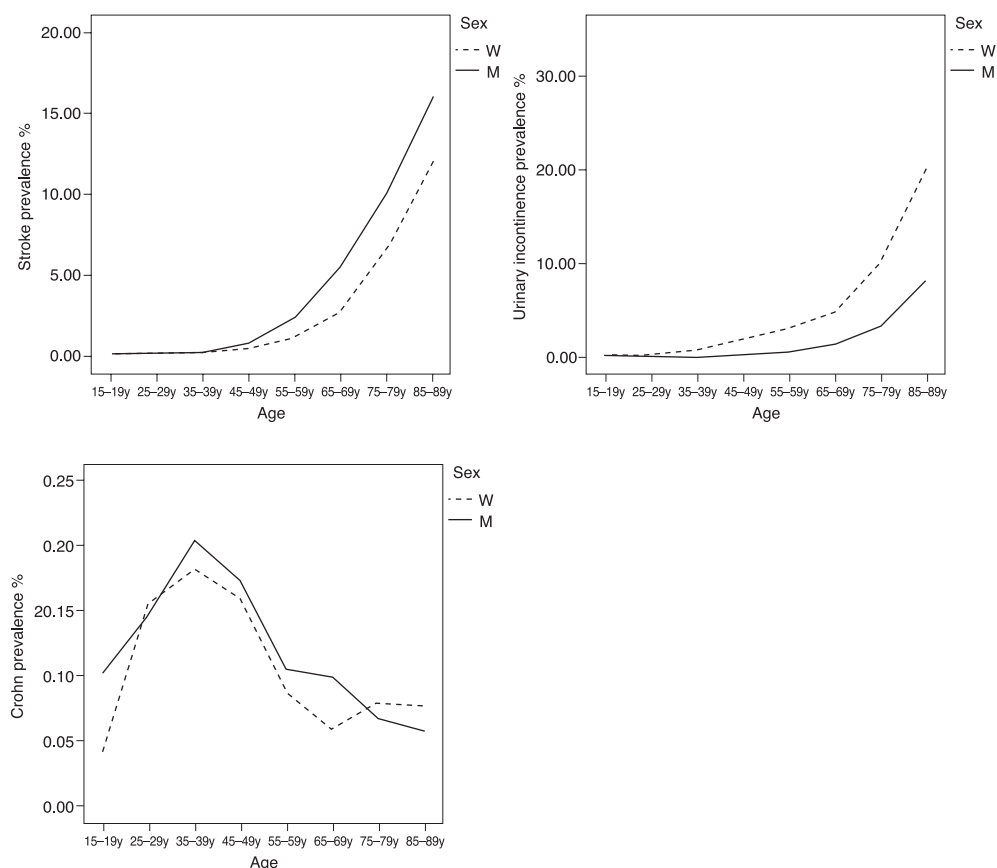


Figure 2 continued

Implications of the findings

Several databases that are similar to SIDIAP in terms of the information collected, data sources and primary care clinical setting are currently being used for cohort, case-control and other study designs. However, most of these databases contain information based on Northern European or North American populations, whereas SIDIAP provides similar clinical variables for Southern Europe. It is this differential characteristic that makes our data particularly interesting. With the addition of this database it will be possible to compare the epidemiology of numerous conditions in Southern and Northern European populations.

The fact that these databases can provide large sample sizes at a comparatively low cost and that they permit long follow-up periods without directly requiring the participation of the subjects, whilst minimising biases such as the Healthy Worker and the Hawthorne effects, has made them especially interesting for public health research. Good examples of applications of database-based studies are the recently published predictive tools made available to clinicians to help estimate the absolute risk of fragility fractures

(QFracture)¹⁹ and of cardiovascular events (QRisk),^{20,21} which have been modelled using QResearch Database data.²²

Limitations

Although rate comparison is a very good approach to ascertain the completeness of electronic records, the main limitation of this study is that we cannot provide an external source of information to allow individualised comparison of the information recorded in our data, because the reference prevalence rates found in the literature may not be correct. Accurate case definition is essential for the reliable reporting of the prevalence of a condition²³ and, as a result, the accuracy of coding cannot be guaranteed at an individual level. However, the fact that the descriptive epidemiology overlaps the known patterns for each of the conditions studied supports the validity of the data. A variety of gold standards have been used, and completeness and validity can only be inferred in relation to the quality of the gold standard used (paper information practice, questionnaires sent to GPs, linkage to external registers,

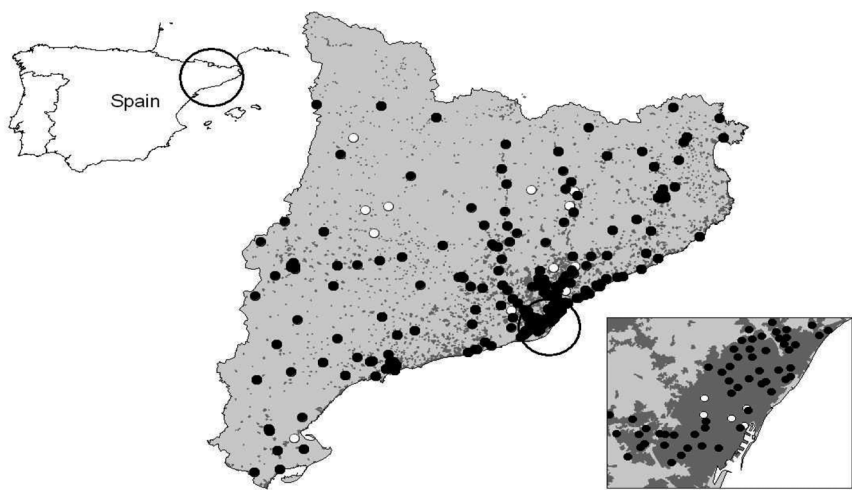
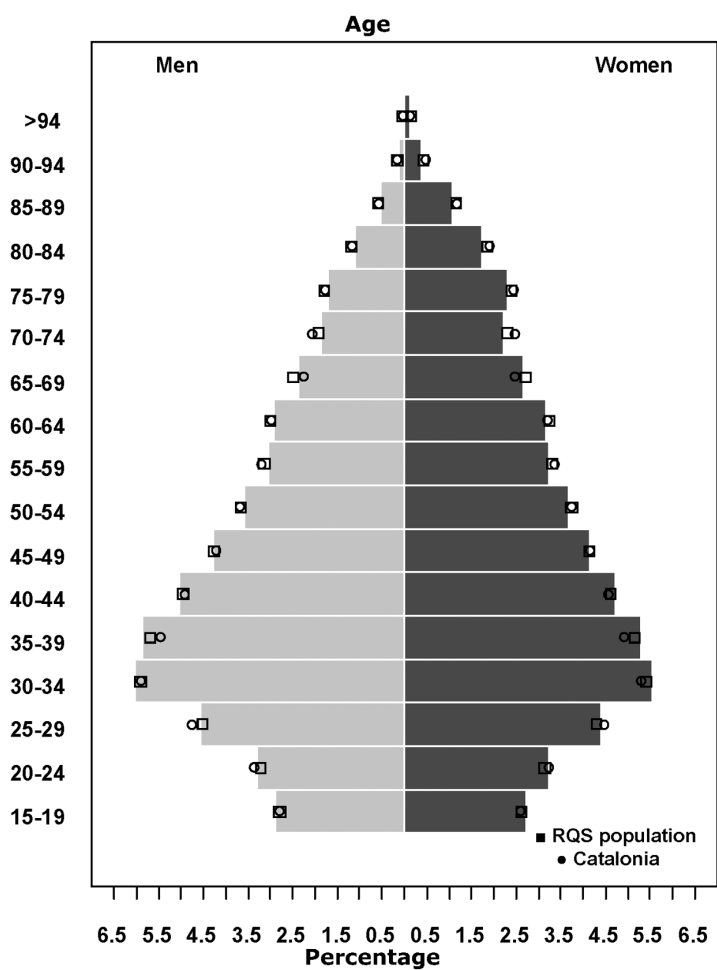


Figure 3(a) and (b) Age-sex population structure and geographical representativeness

hospital discharge databases, reliable cohort studies).²⁴ Techniques like data quality probes to develop internal diagnostic algorithms to identify cases could avoid the problem of the misclassification of diagnostic codes and provide a valuable method for monitoring data accuracy.^{25,26} Moreover, procedures such as participating feedback and audits have shown their usefulness in improving data quality.²⁷ Besides, control charts and cumulative-sum charts can be a good approach for monitoring the cumulative performance of recorded medical information over time.²⁸

In addition, numerous SIDIAP-based projects aim to validate certain conditions using external databases (hospital discharge database, mortality register, etc.) in a similar way as to validations that have been made using the GPRD.^{29,30}

The under-recording we have observed in SIDIAP has also been found to be a common weakness in similar databases. This could lead to a random misclassification error, and therefore to a reduction in statistical power.

Conclusions

SIDIAP contains information on the majority of the population of Catalonia, and is highly representative of the whole region in terms of geographical, age and sex distributions. As we have previously described, more than two thirds of the population of Spain see their GP at least once a year.³¹

Because the information contained in SIDIAP is collected by health professionals during routine visits, it provides a good source of population-based data and reliably reproduces the actual conditions of clinical practice in our setting.

ACKNOWLEDGEMENTS

The authors wish to thank Anna Ponjoan for geographical analysis. This study was supported by the Preventive Services and Health Promotion Network (rediAPP), funded by ISCiii-RETICS (RD06/0018). The funder had no role in the study design, collection, analysis and interpretation of data, writing of the manuscript and decision to submit for publication.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

- 1 Lawrenson R, Williams T and Farmer R. Clinical information for research: the use of general practice databases. *Journal of Public Health Medicine* 1999;21(3):299–304.
- 2 SIDIAP Database (Internet). Available from www.idiapgol.org/sidiap.php
- 3 General Practice Research Database (Internet). Available from www.gprd.com/
- 4 Dietlein G and Schröder-Bernhardi D. Use of the Mediplus patient database in healthcare research. *International Journal of Clinical Pharmacology and Therapeutics* 2002;40(3):130–3.
- 5 Carey IM, Cook DG, De Wilde S et al. Developing a large electronic primary care database (Doctors' Independent Network) for research. *International Journal of Medical Informatics* 2004;73(5):443–53.
- 6 Weiner MG, Lyman JA, Murphy S et al. Electronic Health Records: high-quality electronic data for higher-quality clinical research. *Informatics in Primary Care* 2007;15:121–7.
- 7 Jick H, Jick SS and Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *BMJ* 1991; 302(6779):766–8.
- 8 Jick H, Terris B, Derby LE et al. Further validation of information recorded on a general practitioner based computerized data resource in the United Kingdom. *Pharmacoepidemiology and Drug Safety* 1992;1:347–9.
- 9 Hollowell J. The General Practice Research Database: quality of morbidity data. *Population Trends* 1997; (87):36–40.
- 10 Langley TE, Szatkowski L, Gibson J et al. Validation of The Health Improvement Network (THIN) primary care database for monitoring prescriptions for smoking cessation medications. *Pharmacoepidemiology and Drug Safety* 2010;19(6):586–90.
- 11 Quesada M. The EMMA Project. Atherosclerotic Disease Monitoring Device. IDIAP Jordi Gol Scientific Meeting; 2010. Barcelona, Spain.
- 12 Candel FJ, Matesanz M, Cogolludo F et al. Prevalence of atrial fibrillation and related factors in a population in the centre Madrid. *Anales de Medicina Interna* 2004;21(10):477–82.
- 13 National Health Survey 2006. Available from: www.msps.es/estadEstudios/estadisticas/encuestNacional/ encuesta2006.html
- 14 Basterra-Gortari F and Martinez-Gonzalez M. Comparison of the prevalences of obesity between autonomous regions. *Medicina Clinica (Barcelona)* 2007; 129(12):477–9.
- 15 Tizon JL, Ferrando J, Pares A et al. Schizophrenic disorders in primary care mental health. *Aten Primaria* 2007;39(3):119–24; discussion 125–6.
- 16 Ramos R, Quesada M, Solanas P et al. Prevalence of symptomatic and asymptomatic peripheral arterial disease and the value of the ankle-brachial index to stratify cardiovascular risk. *European Journal of Vascular and Endovascular Surgery* 2009;38(3):305–11.

- 17 Thakar R, Stanton S. Regular review: management of urinary incontinence in women. *BMJ* 2000;321(7272):1326–31.
- 18 Costas Armada P, Garcia-Mayor R *et al.* Prevalence and incidence of Crohn's disease in Galicia. *Medicina Clinica (Barcelona)* 2008;130(18):715.
- 19 Hippisley-Cox J and Coupland C. Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. *BMJ* 2009;339:b4229.
- 20 Hippisley-Cox J, Coupland C, Vinogradova Y *et al.* Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ* 2007;335(7611):136.
- 21 Savill P. Evidence mounts in favour of QRisk. *Practitioner* 2009;253(1721):7.
- 22 Hippisley-Cox J, Stables D and Pringle M. QREsearch: a new general practice database for research. *Informatics in Primary Care* 2004;12(1):49–50.
- 23 de Lusignan, Tomson C, Harris K *et al.* Creatinine fluctuation has a greater effect than the formula to estimate glomerular filtration rate on the prevalence of chronic kidney disease. *Nephron Clinical Practice* 2011; 117:c213–24.
- 24 Jordan K, Porcheret M and Croft P. Quality of morbidity coding in general practice computerized medical records: a systematic review. *Family Practice* 2004;(21): 396–412.
- 25 Brown P and Warmington V. Info-tsunami: surviving the storm with data quality probes. *Informatics in Primary Care* 2003;11:229–37.
- 26 de Lusignan S, Khunti K, Belsey J *et al.* A method of identifying and correcting miscoding, misclassification and misdiagnosis in diabetes: a pilot and validation study of routinely collected data. *Diabetic Medicine* 2010;(27):203–9.
- 27 de Lusignan S. Using feedback to raise the quality of primary care computer data: a literature review. *Studies in Health Technology and Informatics* 2005;116:593–8.
- 28 Noyez L. Control charts, cumsum techniques and funnel plots. A review of methods for monitoring performance in health care. *Interactive Cardiovascular and Thoracic Surgery* 2009;9:494–9.
- 29 Herrett E, Thomas SL, Schoonen WM *et al.* Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *British Journal of Clinical Pharmacology* 2010;69(1):4–14.
- 30 Khan NF, Harrison SE and Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *British Journal of General Practice* 2010;60(572):e128–36.
- 31 0_Pla_de_salut_de_Catalunya_a_l_horitzo2010_1a part_tot.pdf (Objecte application/pdf) [Internet]. Available from: www20.gencat.cat/docs/pla-salut/La_salut/arxius_documents/0_Pla_de_salut_de_Catalunya_a_l_horitzo2010_1a%20part_tot.pdf (accessed 9 March 2011).

ADDRESS FOR CORRESPONDENCE

M^a del Mar García-Gil
 Institut d'Investigació en Atenció Primària (IDIAP-Jordi Gol)
 C) Salvador Maluquer 11
 17002-Girona
 Spain
 Tel: +34 972 487968
 Fax: +34 972 214100
 Email: mgarciagil.girona.ics@gencat.cat

Accepted December 2011